

Operationally-Motivated Uncertainty Relations for Joint Measurability and the Error-Disturbance Tradeoff

Joseph M. Renes and Volkher B. Scholz

Institute for Theoretical Physics, ETH Zurich, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland

We derive new Heisenberg-type uncertainty relations for both joint measurability and the error-disturbance tradeoff for arbitrary observables of finite-dimensional systems. The relations are formulated in terms of a directly operational quantity, namely the probability of distinguishing the actual operation of a device from its hypothetical ideal, by any possible testing procedure whatsoever. Moreover, they may be directly applied in information processing settings, for example to infer that devices which can faithfully transmit information regarding one observable do not leak any information about conjugate observables to the environment. Though intuitively apparent from Heisenberg's original arguments, only more limited versions of this statement have previously been formalized.

I. INTRODUCTION

It is no overstatement to say that the uncertainty principle is a cornerstone of our understanding of quantum mechanics, clearly marking the departure of quantum physics from the world of classical physics. Heisenberg's original formulation in 1927 mentions two facets to the principle. The first restricts the joint measurability of observables, stating that noncommuting observables can only be simultaneously determined with a characteristic amount of indeterminacy [1, p. 172] (see [2, p. 62] for an English translation). The second describes an error-disturbance tradeoff, noting that the more precise a measurement of one observable is made, the greater the disturbance to noncommuting observables [1, p. 175] ([2, p. 64]). The two are of course closely related, and Heisenberg argues for the former on the basis of the latter.

Precise formal statements corresponding to these two facets of the uncertainty principle were constructed only much later, due to the lack of precise mathematical descriptions of measurement and the measurement process in quantum mechanics. Here we must be careful to draw a distinction between statements addressing Heisenberg's original notions of uncertainty from those, like the standard Robertson uncertainty relation [3], which address the impossibility of finding a quantum state with well-defined values for noncommuting observables. Joint measurability has a longer history, going back at least to the seminal work of Arthurs and Kelly [4] and continuing in [5–23]. Quantitative error-disturbance relations themselves have only been formulated relatively recently, going back at least to Braginsky and Khalili [24, Chap. 5] and continuing in [16, 25–31].

One motivation for finding formal uncertainty relations is to delineate the scope and validity of the uncertainty principle. Understanding the fundamental principles of any physical theory is an important endeavor, perhaps doubly so for quantum theory, whose conception of Nature differs so drastically from that used in classical mechanics, not to speak of everyday experience. Our moti-

vation in this article is more operational, however. Here we are interested in finding Heisenberg-type uncertainty relations for joint measurability and error-disturbance that are useful for characterizing and analyzing (quantum) information processing tasks and are formulated in terms of quantities which are immediately relevant in such settings. Indeed, entropic uncertainty relations addressing state preparation [32–34] have already been used to this end, in ensuring the security of quantum key distribution [35, 36]. The foundational and operational motivations are not completely distinct, as concrete information processing settings challenge us to find specific formalizations of the uncertainty principle.

In this article we take a directly operational approach by quantifying error and disturbance in terms of the probability that the actual behavior of the apparatus can be distinguished from a relevant hypothetical behavior, in any experiment whatsoever. We find new uncertainty relations for both joint measurability (Figure 1; Theorem 1) and the error-disturbance tradeoff (Figure 2; Theorem 2) of two arbitrary observables of discrete quantum systems. Our relations address the characteristics of measurement devices themselves, as opposed to entire experimental setups, and can be used in the analysis of quantum information processing tasks. Ultimately, all uncertainty relations spring from the same source, the requirement that the measurement process itself be treated as dynamical process according to the laws of quantum mechanics. The relations presented here are both relatively simple consequences of a basic structure theorem on quantum dynamics, the continuity of the Stinespring representation [37, 38].

We quantify the error made by an apparatus in measuring an observable by the extent to which the apparatus can be distinguished from the *ideal measurement* in any possible experiment. Our uncertainty relation for joint measurability then relates the errors for each observable to a measure of the observables' incompatibility and implies both errors cannot simultaneously be small when the observables are incompatible. On the other hand, we

quantify disturbance to an observable by how well the apparatus mimics one that produces a *fixed output* when acting on states having well-defined values (eigenstates) of that observable. Our error-disturbance tradeoff then relates the error associated with measurement of one observable to the disturbance caused to the other and again implies that both cannot be small when the observables are incompatible.

As mentioned above, entropic state-preparation uncertainty relations have proven useful in establishing the security of quantum key distribution. Our error-disturbance relation allows us to make a stronger statement, one useful in more general cryptographic scenarios beyond creation of secret keys: If a quantum system is subject to any kind of interaction with some external degrees of freedom which nevertheless still allows an experimenter to perform an approximately faithful measurement of a given observable, then the interaction approximately leaks no information about inputs which are eigenstates of a complementary observable. Thus, by appropriately examining the quantum state before and after the interaction, we can infer whether or not any information about the second observable has leaked to the external degrees of freedom. This notion is already present in Heisenberg's original arguments on an intuitive level, but directly operational versions have only been previously formalized for special cases. It can be used to construct leakage-resilient classical computers from fault-tolerant quantum computers [39].

We have organized our results as follows. In the next section we define the distinguishability quantity and provide some background to the mathematical setting of the problem. We then present the joint-measurability and error-disturbance relations, whose proofs are deferred to the Methods section, Sec. V. In Sec. III we discuss the applications to quantum information processing in more detail. Finally, we conclude in Sec. IV with a discussion of open questions raised by this work, in particular how our results could be extended to continuous-variable systems, and a comparison of our results with previous work.

II. MAIN RESULTS

A. Background

When working in the Schrödinger picture, any apparatus is described in the formalism of quantum theory by a completely positive, trace preserving operation, or *quantum channel* [40, 41]. The channel \mathcal{E} maps states in the input state space, $\mathcal{S}(\mathcal{H}_A)$, to states in the output state space, $\mathcal{S}(\mathcal{H}_B)$. Here \mathcal{H}_A and \mathcal{H}_B are Hilbert spaces and $\mathcal{S}(\mathcal{H})$ the set of bounded operators acting on \mathcal{H} and having unit trace.

According to the Stinespring representation theorem [42, 43], any quantum channel $\mathcal{E} : \mathcal{S}(\mathcal{H}_A) \rightarrow \mathcal{S}(\mathcal{H}_B)$

can be expressed in terms of an isometry $V : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$ involving an additional system \mathcal{H}_E as

$$\mathcal{E}(\varrho) = \text{Tr}_E[V\varrho V^\dagger], \quad (1)$$

for any $\varrho \in \mathcal{S}(\mathcal{H}_A)$. The extra system can be regarded as the *environment* of the channel action, the additional degrees of freedom required to describe the dynamics of the pair by the Heisenberg (or Schrödinger) equation. The isometry is however not unique, but all possible Stinespring isometries are related by further isometries involving only the environmental degrees of freedom.

Any particular Stinespring isometry naturally induces another channel, the complementary channel of \mathcal{E} , denoted by \mathcal{E}^\sharp , which maps $\mathcal{S}(\mathcal{H}_A)$ to $\mathcal{S}(\mathcal{H}_E)$ according to

$$\mathcal{E}^\sharp(\varrho) = \text{Tr}_B[V\varrho V^\dagger]. \quad (2)$$

The main technical ingredient required for our results is the continuity of the Stinespring representation [37, 38]. This states that channels which are close (as measured by a particular norm) have Stinespring isometries which are also close. For the formal statement, see Theorem 3.

The error and disturbance measures used here are formulated in terms of the probability $p_{\text{dist}}(\mathcal{E}, \mathcal{E}')$ that one can distinguish the operation of one apparatus \mathcal{E} from another \mathcal{E}' in any test whatsoever, when the two are chosen with equal *a priori* probability. Since this probability ranges from $\frac{1}{2}$ (we can always just make a random guess) to 1, it is more convenient to consider the distinguishability measure

$$\delta(\mathcal{E}, \mathcal{E}') := 2p_{\text{dist}}(\mathcal{E}, \mathcal{E}') - 1, \quad (3)$$

which ranges from zero (completely indistinguishable) to one (completely distinguishable). Fortunately, the distinguishability is directly related to the norm used in the continuity of the Stinespring representation (see (14)). Nonetheless, the operational definition in terms of p_{dist} is sufficient to state our uncertainty relations, and we defer the more detailed presentation of $\delta(\mathcal{E}, \mathcal{E}')$ to Sec. V.

B. Joint Measurability

Let us now consider the question of joint measurability in more detail. As depicted in Figure 1, joint measurability of two observables X and Z is naturally concerned with how well a single apparatus $\mathcal{A}_{X,Z}$ can simultaneously approximate both *ideal measurements*, call them \mathcal{Q}_X and \mathcal{Q}_Z . Any such device has of course two classical outputs, one for each observable, which we denote by R_X and R_Z . The actual measurement \mathcal{M}_X of X made by the apparatus only takes the R_X outcome into account, and similarly for \mathcal{M}_Z . Then, we are specifically interested in the two types of *error* inherent to the apparatus,

$$\varepsilon_X(\mathcal{A}_{X,Z}) := \delta(\mathcal{M}_X, \mathcal{Q}_X) \quad \text{and} \quad (4)$$

$$\varepsilon_Z(\mathcal{A}_{X,Z}) := \delta(\mathcal{M}_Z, \mathcal{Q}_Z). \quad (5)$$

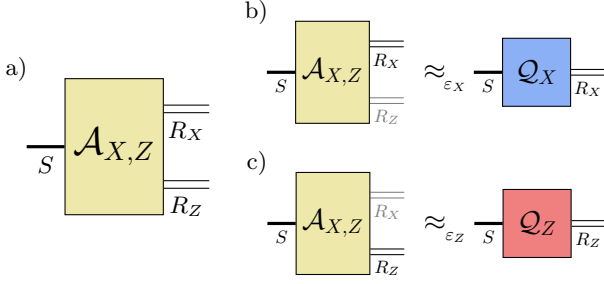


FIG. 1. a) An apparatus $\mathcal{A}_{X,Z}$ designed to jointly measure two observables X and Z on a system S . It produces two results, the classical registers R_X and R_Z . Ignoring either outcome amounts to nonideal measurement \mathcal{M}_X of X and \mathcal{M}_Z of Z , shown in b) and c). The extent to which \mathcal{M}_X and \mathcal{M}_Z fail to simulate ideal measurement \mathcal{Q}_X of the observable X is denoted by the error ϵ_X (defined in (4)) and similarly ϵ_Z for the observable Z . The two errors are constrained by the joint measurability uncertainty relation (7).

We expect that, for incompatible or complementary observables, these quantities cannot both be small. In finite dimensions, we may quantify the complementarity of X and Z in terms of their eigenstates $|\varphi_x\rangle$ and $|\vartheta_z\rangle$, as follows. Letting $r(X; Z) := \frac{1}{\sqrt{2}} (1 - \min_x \max_z |\langle \varphi_x | \vartheta_z \rangle|^2)$, the measure of complementarity is

$$c_1(X, Z) := \max\{r(X; Z), r(Z; X)\}. \quad (6)$$

Then we have the following uncertainty relation,

Theorem 1 (Joint Measurability). *For any apparatus $\mathcal{A}_{X,Z}$ which attempts to jointly measure two finite-dimensional observables X and Z ,*

$$\epsilon_X(\mathcal{A}_{X,Z})^{\frac{1}{2}} + \epsilon_Z(\mathcal{A}_{X,Z})^{\frac{1}{2}} \geq c_1(X, Z). \quad (7)$$

The full proof is given in the Methods section, but we can sketch the main idea here. Since \mathcal{M}_X and \mathcal{M}_Z are defined from the same apparatus, they share a Stinespring isometry, say V . This isometry is close to appropriate isometries W_X and W_Z for \mathcal{Q}_X and \mathcal{Q}_Z as measured by ϵ_X and ϵ_Z , respectively. By the triangle inequality for the isometry distance, we now have a relation for the distance between W_X and W_Z , which can be evaluated by making use of properties of the ideal measurements.

C. Error-Disturbance Tradeoff

Next we turn to the tradeoff between the approximation error of a given apparatus \mathcal{A}_X for measuring observable X and the disturbance caused to the observable Z . The setup is depicted in Figure 2. Again \mathcal{A}_X produces the classical result in R_X , and the approximation error

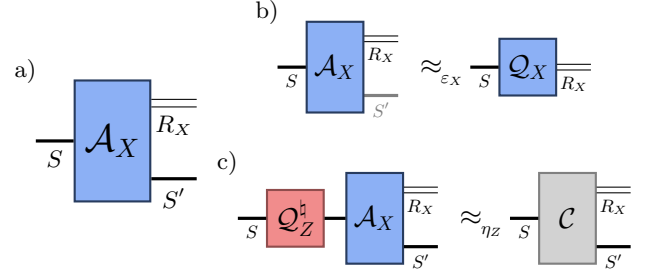


FIG. 2. a) An apparatus designed to extract information about the observable X on a system S . The device produces two outputs: R_X , the classical register carrying information about X , and S' , the transformed quantum system. b) The error of the X measurement is quantified by ϵ_X , the extent to which \mathcal{A}_X approximates an ideal measurement \mathcal{Q}_X . c) The disturbance η_Z is quantified by how well the compound channel $\mathcal{A}_X \circ \mathcal{Q}_Z^1$ approximates a channel \mathcal{C} with a constant output (see (8)); here \mathcal{Q}_Z^1 is an ideal non-selective measurement of Z . The error and disturbance are constrained by the uncertainty relation (10).

$\epsilon_X(\mathcal{A}_X)$ is precisely the same as defined in the previous section. Now we are also interested in the system S' after the action of \mathcal{A}_X , in particular the observable Z .

One measure of disturbance to Z , natural in the Heisenberg picture where the apparatus changes observables on the system, not its state, is how closely the apparatus transforms Z into (some multiple of) the identity operator. This way, any measurement of Z after the action of the apparatus has nothing whatsoever to do with any properties of Z which might have been present beforehand. But here we are after more: The disturbance to Z should also hold conditional on the measurement outcome in R_X . That is, it should not be possible to perform some subsequent “recovery” operation conditional on the measurement outcome which restores the Z observable. This stronger notion of disturbance was used recently in [29].

To formulate a definition of disturbance that addresses this issue yet does explicitly include an optimization over recovery maps requires a little care, particularly in the Heisenberg picture. We first give the logic leading up to our definition for finite-dimensional systems in the Schrödinger picture, and then remark on a natural interpretation in the Heisenberg picture. We give both arguments, as the use of the Heisenberg picture is more convenient when considering infinite-dimensional systems.

For finite-dimensional systems, disturbance to Z in the Schrödinger picture amounts to its eigenstates all being mapped to a fixed output. In the worst case, this is true even when conditioning on the classical outcome of the \mathcal{A}_X apparatus. Therefore, our measure of disturbance is how well the action of \mathcal{A}_X approximates a channel with a constant output, when both are input with eigenstates of Z . To ensure that all inputs to \mathcal{A}_X are diagonal in the Z basis, we may first perform the ideal non-selective

measurement \mathcal{Q}_Z^{\sharp} , which measures the state in the Z basis and discards the result. The post-measurement state is necessarily diagonal in the Z basis, and the map \mathcal{Q}_Z^{\sharp} is a particular complement of the measurement \mathcal{Q}_Z . Therefore, the disturbance is large if the map $\mathcal{A}_X \circ \mathcal{Q}_Z^{\sharp}$ is close to a map \mathcal{C} which has constant output, say σ , for any input state ϱ . We are thus led to a disturbance measure of the form

$$\eta_Z(\mathcal{A}_X) := 1 - \min_{\mathcal{C}} \delta(\mathcal{A}_X \circ \mathcal{Q}_Z^{\sharp}, \mathcal{C}), \quad (8)$$

since a better approximation means greater disturbance.

To motivate this definition in the Heisenberg picture, notice that the non-selective measurement \mathcal{Q}_Z^{\sharp} has no effect on the Z observable itself. Then, to the extent that X and Z are incompatible, \mathcal{Q}_Z^{\sharp} followed by \mathcal{A}_X should completely scramble *all* observables of the system. Indeed, this behavior is measured by (8), since the action of \mathcal{C} on observables is given by $\mathcal{C}'(A) = \text{Tr}[A\sigma]\mathbb{1}$ for the same $\sigma \in \mathcal{S}(\mathcal{H})$ and $A \in \mathcal{B}(\mathcal{H})$, where $\mathcal{B}(\mathcal{H})$ is the set of bounded operators.

As with joint measurement, we expect that both $\varepsilon_X(\mathcal{A}_X)$ and $\eta_Z(\mathcal{A}_X)$ cannot both be small if X and Z are incompatible. For finite-dimensional observables we again measure complementarity in terms of the eigenvectors, but this time by the function

$$c_2(X; Z) := 1 - \max_z \sum_x \left\{ \frac{1}{d} - |\langle \varphi_x | \vartheta_z \rangle|^2 \right\}_+, \quad (9)$$

where $\{x\}_+ = \max\{x, 0\}$ and $d = \dim(\mathcal{H}_S)$. Then we have the following uncertainty relation,

Theorem 2 (Error-Disturbance Tradeoff). *For finite-dimensional observables X and Z , any apparatus \mathcal{A}_X which attempts to gain information about observable X satisfies*

$$\sqrt{2}\varepsilon_X(\mathcal{A}_X)^{\frac{1}{2}} + \eta_Z(\mathcal{A}_X) \geq c_2(X; Z). \quad (10)$$

Again we give a brief sketch of the proof, which is detailed in the Methods section. The first step is to establish an intermediate result, which states any complement of a quantum channel which is close to a measurement \mathcal{Q}_X is itself close to the same measurement, possibly followed by preparation of a new quantum state conditioned on the measurement outcome. Then we consider the channel formed by preceding such a “measure-prepare” channel with the non-selective measurement in the Z basis, \mathcal{Q}_Z^{\sharp} . Finally, the indistinguishability of the resulting joint channel from a constant-output channel turns on how close $\mathcal{Q}_X \circ \mathcal{Q}_Z^{\sharp}$ is to a measurement with a fixed output distribution.

The error-disturbance bound implies that when the error in X is small, then the disturbance to Z must be large relative to $c_2(X; Z)$. However, the opposite conclusion—low disturbance implies high error—does not follow from

the bound, for two reasons. First, the disturbance quantity never quite reaches zero, since it is never possible to perfectly distinguish $\mathcal{A}_X \circ \mathcal{Q}_Z^{\sharp}$ from a constant-output map \mathcal{C} . Second, even if η_Z were zero and the observables conjugate so that $c_2(X; Z) = 1$, ε_X would still only necessarily be at least $\frac{1}{2}$.

III. APPLICATIONS IN QUANTUM INFORMATION PROCESSING

A useful tool in the construction of quantum information processing protocols is the link between reliable transmission of X eigenstates through a channel \mathcal{N} and Z eigenstates through its complement \mathcal{N}^{\sharp} , particularly when the observables X and Z are maximally complementary, i.e. $|\langle \varphi_x | \vartheta_z \rangle|^2 = \frac{1}{d}$ for all x, z . Due to the uncertainty principle, we expect that a channel cannot reliably transmit the bases to different outputs, since this would provide a means to simultaneously measure X and Z . This link has been used by Shor and Preskill to prove the security of quantum key distribution [44] and by Devetak to determine the quantum channel capacity [45]. Entropic state-preparation uncertainty relations from [32, 33] can be used to understand both results, as shown in [46, 47].

However, the above approach has the serious drawback that it can only be used in cases where the specific X -basis transmission over \mathcal{N} and Z -basis transmission over \mathcal{N}^{\sharp} are in some sense compatible and not *counterfactual*; because the argument relies on a state-dependent uncertainty principle, both scenarios must be compatible with the same quantum state. Fortunately, this can be done for both QKD security and quantum capacity, because at issue is whether X -basis (Z -basis) transmission is reliable (unreliable) on average *when the states are selected uniformly at random*. Choosing among either basis states at random is compatible with a random measurement in either basis of half of a maximally-entangled state, and so both X and Z basis scenarios are indeed compatible. The same restriction to choosing input states uniformly appears in the recent result of [29], as it also ultimately relies on a state-preparation uncertainty relation.

Using Theorem 2 we can extend the method above to counterfactual uses of arbitrary channels \mathcal{N} , in the following sense: If acting with the channel \mathcal{N} does not substantially affect the possibility of performing an X measurement, then Z -basis inputs to \mathcal{N}^{\sharp} result in an essentially constant output. More concretely, we have

Corollary 1. *Given a channel \mathcal{N} and complementary channel \mathcal{N}^{\sharp} , suppose that there exists a measurement Λ_X such that $\delta(\mathcal{Q}_X, \Lambda_X \circ \mathcal{N}) \leq \varepsilon$. Then there exists a constant channel \mathcal{C} such that $\delta(\mathcal{N}^{\sharp} \circ \mathcal{Q}_Z^{\sharp}, \mathcal{C}) \leq 2\sqrt{\varepsilon} + 1 - c_2(X; Z)$. For maximally complementary X and Z , $\delta(\mathcal{N}^{\sharp} \circ \mathcal{Q}_Z^{\sharp}, \mathcal{C}) \leq 2\sqrt{\varepsilon}$.*

Proof. Let $V_{\mathcal{N}}$ be the Stinespring dilation of \mathcal{N} such that \mathcal{N}^{\sharp} is the complementary channel and define $\mathcal{A}_X = \Lambda_X \circ$

V_N . For \mathcal{C} the optimal choice in the definition of $\eta_Z(\mathcal{A}_X)$, (10) implies $\delta(\mathcal{A}_X \circ \mathcal{Q}_Z^h, \mathcal{C}) \leq 2\sqrt{\varepsilon} + 1 - c_2(X; Z)$. Since \mathcal{N}^\sharp is obtained from \mathcal{A}_X by ignoring the Λ_X measurement result, $\delta(\mathcal{N}^\sharp \circ \mathcal{Q}_Z^h, \mathcal{C}) \leq \delta(\mathcal{A}_X \circ \mathcal{Q}_Z^h, \mathcal{C})$. \square

This formulation is important because in more general cryptographic and communication scenarios we are interested in the worst-case behavior of the protocol, not the average case under some particular probability distribution. For instance, in [39] the goal is to construct a classical computer resilient to leakage of Z -basis information by establishing that reliable X basis measurement is possible despite the interference of the eavesdropper. However, such an X measurement is entirely counterfactual and cannot be reconciled with the actual Z -basis usage, as the Z -basis states will be chosen *deterministically* in the classical computer.

IV. OUTLOOK

A number of open questions present themselves, beyond an extension of our results to continuous variable systems, which is discussed at length below. First, it would be interesting to examine the optimality conditions of the semidefinite program used in the proofs of both relations to see if the bounds presented here could be improved. Both are somewhat weak in certain extreme cases: In the joint-measurability relation (7) even if one measurement is perfect the bound on the error of the other is only at least one half, while from the error-disturbance relation 10 one cannot conclude that low disturbance implies high error. One could also examine the tightness of either relation (at least numerically) for specific measurement devices, for instance the measurement used in the experimental tests [48, 49] of Ozawa's error-disturbance relation [26]. Finally, one could also derive bounds on joint-measurability from the error-disturbance tradeoff itself and see how it compares with (7).

A. Extension to Continuous Variables

Starting from Heisenberg's seminal paper, clearly the most well-studied uncertainty relations involve the conjugate pair position and momentum. Hence it is desirable to extend our results to this setting as well. The two main technical tools used establish Theorems 1 and 2 are the continuity of the Stinespring representation for completely positive maps (Theorem 3, as well as formulation of the completely bounded norm as a semidefinite program involving the Choi representative of a map (Equations 16 and 17). Both results can be carried over to the case of infinite dimensional systems. The continuity theorem remains valid as stated in Theorem 3, see [37]. In addition, since we are concerned with the distinguishability of measurements, which are channels that destroy

all entanglement (known as entanglement-breaking channels), the concept of Choi representatives carries over as well [50, Theorem 3].

Before applying these two tools in the infinite-dimensional setting, we must first ensure that the definitions of error and disturbance are sensible. Recall that both rest on the probability of distinguishing between two channels, maximized over all input states and observables. In infinite dimensions this includes observables of arbitrary *precision*; any operationally valid distinguishability measure should however also take precision into account. For finite dimensions precision is not an issue due to the inherent discreteness of the results (infinite precision is not unphysical, in some sense).

For the case of our error-disturbance relation, the extension to continuous variables is even more mathematically delicate, as there simply is no instrument which implements an ideal position or momentum measurement, including the post-measurement quantum state [6, Theorem 3.3]. Instruments implementing imprecise measurements do however exist. A meaningful extension of our joint measurability result would not need to address this difficulty, since ideal position and momentum measurements themselves do exist, ignoring the post-measurement quantum state.

In the case of distinguishing measurements, as here, the observables in question are so-called test functions, which take values between 0 and 1, generalizing characteristic or indicator functions associated with a given subset of the measurement output space (the real line for position or momentum). One way to limit the precision of test functions would be to ensure they are "smeared out" over some minimum length scale, for instance by composing all test functions with a physical noise channel. Our proof technique could potentially be adapted to include this additional step, since by using a physical noise channel one could still make use of the Stinespring representation.

Another option would be to restrict to slowly-varying functions, that is having a bounded Lipschitz constant, which are thereby insensitive to changes on small length scales. This approach is chosen in the work of Werner [18], and the resulting measure of distinguishability between probability distributions is the Wasserstein metric of order one. Thanks to the theory of optimal transport, there is a nice dual interpretation of the distinguishability measure: It is the cost required to change one distribution into the other, as measured by the distance on the real line. Note that the variational distance, which underlies the completely bounded norm as used here, can be formulated similarly. Now the cost is measured by a different metric which is simply zero if the two values are identical and one otherwise [51]. Hence a possible way to incorporate finite precision limits into our setup would be to require the metric to be only sensitive to differences above some finite minimal length scale. It would be interesting to formalize both approaches and in particular their "completely bounded" versions.

B. Comparison to Previous Results

In the recent work of Busch, Lahti and Werner [21], the authors used the Wasserstein metric of order two, corresponding to the mean squared error, as the underlying distance $D(.,.)$ to measure the closeness of probability distributions. If $\mathcal{M}^Q, \mathcal{M}^P$ are the marginals of some joint measurement of position Q and momentum P , and X_ϱ denotes the distribution coming from applying the measurement X to the state ϱ , their relation reads

$$\sup_{\varrho} D(\mathcal{M}_\varrho^Q, Q_\varrho) \cdot \sup_{\varrho} D(\mathcal{M}_\varrho^P, P_\varrho) \geq c, \quad (11)$$

for some universal constant c . In [23], the authors generalize their results to arbitrary Wasserstein metrics. As in our case, the two distinguishability quantities in (11) are separately maximized over all states, and hence the resulting expression characterizes the goodness of the approximate measurement.

One could instead ask for a “coupled optimization”, a relation of the form

$$\sup_{\varrho} [D(\mathcal{M}_\varrho^Q, Q_\varrho) D(\mathcal{M}_\varrho^P, P_\varrho)] \geq c', \quad (12)$$

for some other constant c' .¹ While this statement certainly tells us that no device can accurately measure both position and momentum for all input states, the bound c' only holds (and can only hold) for the worst possible input state. In contrast, the bounds found in Theorems 1 and 2, as well as in (11) are state-independent in the sense that the bound holds for all states. Indeed, the two approaches are more distinct than the similarities between (11) and (12) would suggest. By optimizing over input states separately, our results and those of [18, 21, 23] are statements about the properties of measurement devices themselves, independent of any particular experimental setup. State-dependent settings capture the behavior of measurement devices in specific experimental setups and must therefore account for the details of the input state.

The same set of authors also studied the case of finite-dimensional systems, in particular qubit systems, again using the Wasserstein metric of order two [22]. Their results for this case are similar, with the product in (11) replaced by a sum. Perhaps most closely related to our results is the recent work by Ipsen [30], who uses the variational distance as the underlying distinguishability measure to derive similar additive uncertainty relations. We note, however, that both [22] and [30] only consider joint measurability and do not consider the change to the state after the approximate measurement is performed,

as it is done in our error-disturbance relation. Furthermore, both base their distinguishability measures on the measurement statistics of the devices alone. But this does not necessarily tell us how distinguishable two devices ultimately are, as we could employ input states entangled with ancilla systems to test them. These two measures can be different [52], even for entanglement-breaking channels [53].

Entropic quantities are another means of comparing two probability distributions, an approach taken recently by Buscemi *et al.* [29] and Coles and Furrer [31] (see also Martens and de Muynck [25]). Both contributions formalize error and disturbance in terms of relative or conditional entropies, and derive their results from entropic uncertainty relations for state preparation which incorporate the effects of quantum entanglement [32, 33]. They differ in the choice of the entropic measure and the choice of the state on which the entropic terms are evaluated. Buscemi *et al.* find state-independent error-disturbance relations involving the von Neumann entropy, evaluated for input states which describe observable eigenstates chosen uniformly at random. As described in Sec. III, the restriction to uniformly-random inputs is significant, and leads to a characterization of the average-case behavior of the device (averaged over the choice of input state), not the worst-case behavior as presented here. Meanwhile, Coles and Furrer make use of general Rényi-type entropies, hence also capturing the worst-case behavior. However, they are after a state-dependent error-disturbance relation which relates the amount of information a measurement device can extract from a state about the results of a *future* measurement of one observable to the amount of disturbance caused to other observable.

An important distinction between both these results and those presented here is the quantity appearing in the uncertainty bound, i.e. the quantification of complementarity of two observables. As both the aforementioned results are based on entropic state-preparation uncertainty relations, they each quantify complementarity by the largest overlap of the eigenstates of the two observables. This bound is trivial should the two observables share an eigenstate. However, a perfect joint measurement is clearly impossible even if the observables share all but two eigenvectors (if they share all but one, they necessarily share all eigenvectors). Both $c_1(X, Z)$ and $c_2(X; Z)$ used here are nontrivial whenever not all eigenvectors are shared between the observables.

ACKNOWLEDGMENTS

Thanks to Omar Fawzi, Fabian Furrer, David Reeb, and Michael Walter for helpful discussions. This work was supported by the by the German Science Foundation (grant CH 843/2-1), the Swiss National Science Foundation (through the National Centre of Competence in Research ‘Quantum Science and Technology’ and grants No. 200020-135048, PP00P2-128455, 20CH21-138799

¹ Such an approach has been advocated by David Reeb (private communication).

(CHIST-ERA project CQC)), by the European Research Council (grants No. 258932 and No. 337603) as well as by the Swiss State Secretariat for Education and Research through COST action MP1006. VBS is supported by an ETH Postdoctoral Fellowship.

V. METHODS

A. Mathematical Setup

A result of Helstrom [54, 55] shows that the distinguishability of two quantum states ϱ_1 and ϱ_2 is precisely their trace distance, $\delta(\varrho_1, \varrho_2) := \frac{1}{2} \|\varrho_1 - \varrho_2\|_1$, where $\|A\|_1 := \text{Tr}[\sqrt{A^\dagger A}]$. In other words, $p = \frac{1}{2}(1 + \delta(\varrho_1, \varrho_2))$. Distinguishability of states can be transferred to that of channels by asking for the most distinguishable states that two channels could produce from a common input. Since this distinguishability can be enhanced for inputs which are entangled with ancillary systems unaffected by the channel itself, one is lead to consider the *diamond norm* of quantum channels [52].

For a channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$, the diamond norm is defined by

$$\|\mathcal{E}\|_\diamond := \sup_{k \geq 1} \sup_{\varrho \in \mathcal{S}(\mathcal{H}_A \otimes \mathbb{C}^k)} \|\mathcal{E} \otimes \mathcal{I}_k(\varrho)\|_1, \quad (13)$$

where \mathcal{I}_k is the identity channel from $\mathcal{B}(\mathbb{C}^k)$ to itself which just reproduces its input. Using the diamond norm we arrive at the following distinguishability measure for quantum channels,

$$\delta(\mathcal{E}_1, \mathcal{E}_2) = \frac{1}{2} \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond. \quad (14)$$

This expression is not closed-form, as an optimization is required to evaluate the diamond norm. However, in finite dimensions the diamond norm can be cast as a convex optimization, specifically as a semidefinite program [56]. This makes numerical approximation tractable and will be analytically useful in the proofs to follow.

Given a Hilbert space \mathcal{H} with basis $\{|k\rangle\}_{k=1}^d$, define $|\Omega\rangle = \sum_{k=1}^d |k\rangle \otimes |k\rangle \in \mathcal{H} \otimes \mathcal{H}$. Then, for any channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$, let \mathcal{C} denote the Choi mapping of \mathcal{E} to a bipartite state,

$$\mathcal{C}(\mathcal{E}) := \mathcal{E} \otimes \mathcal{I}(|\Omega\rangle\langle\Omega|) \in \mathcal{S}(\mathcal{H}_B \otimes \mathcal{H}_A). \quad (15)$$

The diamond norm can then be expressed as

$$\begin{aligned} \frac{1}{2} \|\mathcal{E}\|_\diamond &= \max \text{Tr}[\mathcal{C}(\mathcal{E})Y] \\ \text{s.t. } Y &\leq \mathbb{1}_B \otimes \varrho_A, \\ Y &\geq 0, \\ \varrho &\in \mathcal{S}(\mathcal{H}_A). \end{aligned} \quad (16)$$

This semidefinite program also comes in the dual form

$$\begin{aligned} \frac{1}{2} \|\mathcal{E}\|_\diamond &= \min \|\text{Tr}_B(R)\|_\infty \\ \text{s.t. } R &\geq \mathcal{C}(\mathcal{E}), \\ R &\geq 0 \end{aligned} \quad (17)$$

Here we have used the operator norm, defined by $\|A\|_\infty = \sup_{\psi \in \mathcal{H}} \langle \psi | A | \psi \rangle$. Both forms of the semidefinite program will be useful in the proofs to follow.

Finally, we arrive at the central technical tool required for our uncertainty relations, the continuity of the Stinespring representation for finite-dimensional systems.

Theorem 3 (Stinespring Continuity [38]). *Given two quantum channels $\mathcal{E}_1, \mathcal{E}_2 : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$ with corresponding Stinespring isometries $V_1 : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_{E_1}$ and $V_2 : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_{E_2}$, we have*

$$\min_U \|UV_1 - V_2\|_\infty^2 \leq \|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond \leq 2 \min_U \|UV_1 - V_2\|_\infty,$$

with the minimum taken over all isometries $U : \mathcal{H}_{E_1} \rightarrow \mathcal{H}_{E_2}$.

B. Joint Measurability

A device $\mathcal{A}_{X,Z}$ jointly measuring two observables X, Z on a Hilbert space \mathcal{H}_S can be modeled by an isometry

$$\begin{aligned} V : \mathcal{H}_S &\rightarrow \mathcal{H}_R \otimes \mathcal{H}_{\hat{X}} \otimes \mathcal{H}_{\hat{X}'} \otimes \mathcal{H}_{\hat{Z}} \otimes \mathcal{H}_{\hat{Z}'}, \\ V &= \sum_{xz} M_{xz} \otimes |x\rangle_{\hat{X}} \otimes |x\rangle_{\hat{X}'} \otimes |z\rangle_{\hat{Z}} \otimes |z\rangle_{\hat{Z}'}. \end{aligned} \quad (18)$$

Here, the Hilbert spaces $\mathcal{H}_{\hat{X}}, \mathcal{H}_{\hat{Z}}$ serve to record the (classical) measurement result, and doubling them ensures that no quantum coherence is present in systems \hat{X} or \hat{Z} alone. The spaces have dimension equal to the number of outputs, and x, z label arbitrary bases (quite possibly the same). The operators $M_{x,z} : \mathcal{H}_S \rightarrow \mathcal{H}_R$ are the Kraus operators of $\mathcal{A}_{X,Z}$. The maps $\mathcal{M}_X : \mathcal{S}(\mathcal{H}_S) \rightarrow \mathcal{S}(\mathcal{H}_{\hat{X}})$ and $\mathcal{M}_Z : \mathcal{S}(\mathcal{H}_S) \rightarrow \mathcal{S}(\mathcal{H}_{\hat{Z}})$ describing the measurement of X or Z by the device are determined by

$$\mathcal{M}_X(\varrho) = \text{Tr}_{R\hat{Z}\hat{Z}'}[V\varrho V^\dagger], \quad \text{and} \quad (19)$$

$$\mathcal{M}_Z(\varrho) = \text{Tr}_{R\hat{X}\hat{X}'}[V\varrho V^\dagger]. \quad (20)$$

Particular Stinespring isometries for the ideal measurements $\mathcal{Q}_X, \mathcal{Q}_Z$ are given by

$$\begin{aligned} W_X : \mathcal{H}_S &\rightarrow \mathcal{H}_S \otimes \mathcal{H}_{\hat{X}} \otimes \mathcal{H}_{\hat{X}'}, \\ W_X &= \sum_x Q_X(x) \otimes |x\rangle_{\hat{X}} \otimes |x\rangle_{\hat{X}'}, \quad \text{and} \end{aligned} \quad (21)$$

$$\begin{aligned} W_Z : \mathcal{H}_S &\rightarrow \mathcal{H}_S \otimes \mathcal{H}_{\hat{Z}} \otimes \mathcal{H}_{\hat{Z}'}, \\ W_Z &= \sum_z Q_Z(z) \otimes |z\rangle_{\hat{Z}} \otimes |z\rangle_{\hat{Z}'}, \end{aligned} \quad (22)$$

where $Q_X(x)$ and $Q_Z(z)$ denote the projection operators associated with the observables X or Z . In terms of

the eigenvectors of the respective observables, $Q_X(x) = |\varphi_x\rangle\langle\varphi_x|$ and $Q_Z(z) = |\vartheta_z\rangle\langle\vartheta_z|$.

According to the lower bound in the Stinespring continuity Theorem 3, there exist isometries $U_X : \mathcal{H}_S \rightarrow \mathcal{H}_R \otimes \mathcal{H}_{\hat{X}} \otimes \mathcal{H}_{\hat{X}'}$, and $U_Z : \mathcal{H}_S \rightarrow \mathcal{H}_R \otimes \mathcal{H}_{\hat{Z}} \otimes \mathcal{H}_{\hat{Z}'}$ such that

$$\|V - U_X W_Z\|_\infty^2 \leq \|\mathcal{M}_Z - \mathcal{Q}_Z\|_\diamond, \quad (23)$$

$$\|V - U_Z W_X\|_\infty^2 \leq \|\mathcal{M}_X - \mathcal{Q}_X\|_\diamond. \quad (24)$$

Using the triangle inequality and the definition of error ε , we find

$$\frac{1}{\sqrt{2}} \|U_X W_Z - U_Z W_X\|_\infty \leq \varepsilon_X (\mathcal{A}_{X,Z})^{\frac{1}{2}} + \varepsilon_Z (\mathcal{A}_{X,Z})^{\frac{1}{2}}. \quad (25)$$

In principle, this inequality already gives a bound on the errors ε_X and ε_Z . However, it is implicitly a function of the measurement device, since U_X and U_Z are only characterized by the optimal choice in the Stinespring dilation, which itself turns on the description of the device. For finite-dimensional systems, we can find a bound which holds for all devices as follows.

Proof of Theorem 1. First define the map $\mathcal{E}_Z : \mathcal{B}(\mathcal{H}_S) \rightarrow \mathcal{B}(\mathcal{H}_{\hat{X}})$ by

$$\mathcal{E}_Z(\varrho) := \text{Tr}_{R\hat{Z}\hat{Z}'\hat{X}'}[U_X W_Z \varrho W_Z^\dagger U_X^\dagger], \quad (26)$$

$$= \text{Tr}_{R\hat{X}'}[U_X \mathcal{Q}_Z^\dagger(\varrho) U_X^\dagger]. \quad (27)$$

Here, we have used the map \mathcal{Q}_Z^\dagger , which will be used later in the definition of disturbance; it is sometimes called a “pinch map” and is defined by

$$\mathcal{Q}_Z^\dagger(\varrho) = \sum_z Q_Z(z) \varrho Q_Z(z) \quad (28)$$

$$= \sum_z \langle \vartheta_z | \varrho | \vartheta_z \rangle |\vartheta_z\rangle\langle\vartheta_z|_S. \quad (29)$$

The ideal measurement is $\mathcal{Q}_X(\varrho) = \text{Tr}_{S\hat{X}'}[W_X \varrho W_X^\dagger]$, which can also be expressed as $\mathcal{Q}_X(\varrho) = \text{Tr}_{R\hat{Z}\hat{Z}'\hat{X}'}[U_Z W_X \varrho W_X^\dagger U_Z^\dagger]$. By Stinespring continuity we therefore have

$$\|U_X W_Z - U_Z W_X\|_\infty \geq \frac{1}{2} \|\mathcal{Q}_X - \mathcal{E}_Z\|_\diamond. \quad (30)$$

Now we make use of the primal form of the semidefinite program given in (16). We are free to choose the basis in which $|\Omega\rangle$ is defined, so let us select $\{|\varphi_x\rangle\}$, the eigenbasis of the X observable. Then we find $\mathcal{C}(\mathcal{Q}_X) = \sum_x |x\rangle\langle x|_{\hat{X}} \otimes |\varphi_x\rangle\langle\varphi_x|_S$, while

$$\begin{aligned} \mathcal{C}(\mathcal{E}_Z) &= \sum_{z,y,y'} \langle \varphi_{y'} | Q_Z(z) | \varphi_y \rangle \\ &\quad \times \text{Tr}_{R\hat{X}'}[U_X Q_Z(z) U_X^\dagger] \otimes |\varphi_y\rangle\langle\varphi_{y'}|_S. \end{aligned} \quad (31)$$

Now define $\Lambda_x := U_X^\dagger(\mathbb{1}_R \otimes |x\rangle\langle x|_{\hat{X}} \otimes \mathbb{1}_{\hat{X}'})U_X$ and let $Y = |x\rangle\langle x|_{\hat{X}} \otimes |\varphi_x\rangle\langle\varphi_x|$ for some x . From (16) we get

$$\begin{aligned} &\frac{1}{2} \|\mathcal{Q}_X - \mathcal{E}_Z\|_\diamond \\ &\geq \max_x \left(1 - \sum_z \text{Tr}[Q_Z(z) \Lambda_x] |\langle \vartheta_z | \varphi_x \rangle|^2\right) \end{aligned} \quad (32)$$

$$\geq 1 - \min_x \left(\max_z |\langle \vartheta_z | \varphi_x \rangle|^2 \sum_z \text{Tr}[Q_Z(z) \Lambda_x]\right) \quad (33)$$

$$= 1 - \min_x \left(\max_z |\langle \vartheta_z | \varphi_x \rangle|^2 \text{Tr}[\Lambda_x]\right) \quad (34)$$

$$\geq 1 - \min_x \left(\max_z |\langle \vartheta_z | \varphi_x \rangle|^2\right) \min_x \text{Tr}[\Lambda_x] \quad (35)$$

$$\geq 1 - \min_x \max_z |\langle \vartheta_z | \varphi_x \rangle|^2. \quad (36)$$

The last inequality follows because $\sum_x \Lambda_x = U_X^\dagger U_X = \mathbb{1}_S$, which then implies $\sum_{x=1}^d \text{Tr}[\Lambda_x] = d$ and therefore $\min_x \text{Tr}[\Lambda_x] \leq 1$. \square

C. Error-Disturbance Tradeoff

A device \mathcal{A}_X that measures X on a Hilbert space \mathcal{H}_S and also produces an output state on \mathcal{H}_R can be modelled by an isometry

$$\begin{aligned} V : \mathcal{H}_S &\rightarrow \mathcal{H}_R \otimes \mathcal{H}_E \otimes \mathcal{H}_{\hat{X}} \otimes \mathcal{H}_{\hat{X}'}, \\ V &= \sum_x M_x \otimes |x\rangle_{\hat{X}} \otimes |x\rangle_{\hat{X}'}. \end{aligned} \quad (37)$$

Again the $M_x : \mathcal{H}_S \rightarrow \mathcal{H}_R \otimes \mathcal{H}_E$ are the Kraus operators of the map, while \mathcal{H}_E is an extra system which may be needed to purify the output in R . As with $\mathcal{A}_{X,Z}$, the measurement outcome is recorded in $\mathcal{H}_{\hat{X}}$ (and $\mathcal{H}_{\hat{X}'}$).

Before proceeding to the proof of Theorem 2, we first establish an intermediate result which states that complementary channels of approximate measurements are themselves approximate measurements, possibly followed by state preparation. More precisely, if \mathcal{A}_X approximates the ideal measurement of X , then the output in system R can be simulated by a map which simply prepares a state on \mathcal{H}_R conditional on the result x of the measurement. We call $\mathcal{P} : \mathcal{S}(\mathcal{H}_{\hat{X}}) \rightarrow \mathcal{S}(\mathcal{H}_R \otimes \mathcal{H}_{\hat{X}})$ a conditional preparation channel if it has the action $\mathcal{P}(|x\rangle\langle x|) = \varrho_R^x \otimes |x\rangle\langle x|_{\hat{X}}$, for some states ϱ_R^x . Then, by Stinespring continuity we can show

Theorem 4. *Given a channel $\mathcal{A}_X : \mathcal{S}(\mathcal{H}_S) \rightarrow \mathcal{S}(\mathcal{H}_R \otimes \mathcal{H}_{\hat{X}})$, let $\mathcal{M}_X : \mathcal{S}(\mathcal{H}_S) \rightarrow \mathcal{S}(\mathcal{H}_{\hat{X}})$ be just the output in \hat{X} , i.e. $\mathcal{M}_X(\varrho) = \text{Tr}_{R\hat{X}'}[\mathcal{A}_X(\varrho)]$. If*

$$\|\mathcal{M}_X - \mathcal{Q}_X\|_\diamond \leq \varepsilon, \quad (38)$$

then there exists a conditional state preparation channel $\mathcal{P} : \mathcal{S}(\mathcal{H}_{\hat{X}}) \rightarrow \mathcal{S}(\mathcal{H}_R \otimes \mathcal{H}_{\hat{X}})$ such that

$$\|\mathcal{A}_X - \mathcal{P} \circ \mathcal{Q}_X\|_\diamond \leq 2\sqrt{\varepsilon}. \quad (39)$$

Proof. We can reuse the Stinespring dilation W_X of \mathcal{Q}_X given in (21), while the dilation of \mathcal{A}_X is given by (37). By the lower bound in Theorem 3, the premise above implies

$$\|V - UW_X\|_\infty \leq \sqrt{\varepsilon}, \quad (40)$$

for some isometry $U : \mathcal{H}_S \rightarrow \mathcal{H}_R \otimes \mathcal{H}_E$.

Now, for any $\varrho \in \mathcal{S}(\mathcal{H}_S)$,

$$\mathcal{A}_X(\varrho) = \text{Tr}_{E\hat{X}'}[V\varrho V^\dagger]. \quad (41)$$

Thus, defining the map $\mathcal{E} : \mathcal{S}(\mathcal{H}_S) \rightarrow \mathcal{S}(\mathcal{H}_R \otimes \mathcal{H}_{\hat{X}})$ by

$$\mathcal{E}(\varrho) := \text{Tr}_{E\hat{X}'}[UW_X\varrho W_X^\dagger U^\dagger] \quad (42)$$

$$= \text{Tr}_E[U\text{Tr}_{\hat{X}'}[W_X\varrho W_X^\dagger]U^\dagger]. \quad (43)$$

the upper bound in Theorem 3 implies

$$\|\mathcal{A}_X - \mathcal{E}\|_\diamond \leq 2\sqrt{\varepsilon}. \quad (44)$$

All that remains to show is that $\mathcal{E} = \mathcal{P} \circ \mathcal{Q}_X$ for some conditional preparation channel \mathcal{P} . Using the form of W_X we can express the action of \mathcal{E} as

$$\mathcal{E}(\varrho) = \sum_x \text{Tr}_E[UQ_X(x)\varrho Q_X(x)U^\dagger] \otimes |x\rangle\langle x|_{\hat{X}} \quad (45)$$

$$= \sum_x \langle \varphi_x | \varrho | \varphi_x \rangle \text{Tr}_E[U|\varphi_x\rangle\langle\varphi_x|U^\dagger] \otimes |x\rangle\langle x|_{\hat{X}}. \quad (46)$$

This is a conditional preparation channel \mathcal{P} for $\varrho_R^x = \text{Tr}_E[U|\varphi_x\rangle\langle\varphi_x|U^\dagger]$. \square

Now we can establish the error-disturbance bound.

Proof of Theorem 2. An apparatus \mathcal{A}_X with error $\varepsilon_X(\mathcal{A}_X)$ satisfies the premise of Theorem 4 with $\varepsilon = 2\varepsilon_X(\mathcal{A}_X)$, and therefore

$$\delta(\mathcal{A}_X, \mathcal{P} \circ \mathcal{Q}_X) \leq \sqrt{2\varepsilon_X(\mathcal{A}_X)} \quad (47)$$

for some conditional preparation channel \mathcal{P} . By the triangle inequality, for any map \mathcal{C} we have

$$\begin{aligned} \delta(\mathcal{A}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{C}) &\leq \delta(\mathcal{A}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{P} \circ \mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger) \\ &\quad + \delta(\mathcal{P} \circ \mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{C}). \end{aligned} \quad (48)$$

By monotonicity, the first term on the righthand side is less than $\delta(\mathcal{A}_X, \mathcal{P} \circ \mathcal{Q}_X)$ and therefore

$$\delta(\mathcal{A}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{C}) \leq \sqrt{2\varepsilon_X(\mathcal{A}_X)} + \delta(\mathcal{P} \circ \mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{C}). \quad (49)$$

Next, let $\mathcal{F} : \mathcal{S}(\mathcal{H}_S) \rightarrow \mathcal{H}_{\hat{X}}$ be the channel with action $\mathcal{F}(\varrho) \mapsto \sum_x p_x |x\rangle\langle x|$ for some fixed probability distribution p_x and define $\mathcal{C} = \mathcal{P} \circ \mathcal{F}$. Again by monotonicity, the second term on the righthand side is less than $\delta(\mathcal{C} \circ \mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{F})$, and so we obtain

$$\min_{\mathcal{C}} \delta(\mathcal{A}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{C}) \leq \sqrt{2\varepsilon_X(\mathcal{A}_X)} + \delta(\mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{F}), \quad (50)$$

since we are free to minimize over the maps \mathcal{C} . This expression is equivalent to

$$\sqrt{2\varepsilon_X(\mathcal{A}_X)}^{\frac{1}{2}} + \eta_Z(\mathcal{A}_X) \geq 1 - \delta(\mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{F}). \quad (51)$$

Now we make use of the dual form, Eq. (17), of the semidefinite program for the diamond norm to find an upper bound on $\delta(\mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger, \mathcal{F})$. We first compute $\mathcal{C}(\mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger - \mathcal{F})$ and then make a suitable choice of R . Choosing the basis of $|\Omega\rangle$ to be the Z basis $|\vartheta_z\rangle$, we find that

$$\mathcal{C}(\mathcal{Q}_X \circ \mathcal{Q}_Z^\dagger) = \sum_{xz} |\langle \varphi_x | \vartheta_z \rangle|^2 |x\rangle\langle x|_{\hat{X}} \otimes |\vartheta_z\rangle\langle \vartheta_z|_S. \quad (52)$$

For $\mathcal{C}(\mathcal{F})$ we have simply $\mathcal{C}(\mathcal{F}) = \sum_x p_x |x\rangle\langle x|_{\hat{X}} \otimes \mathbb{1}_S$. Choose $p_x = \frac{1}{d}$ and define

$$R_{\hat{X}S} = \sum_{xz} \left\{ \frac{1}{d} - |\langle \varphi_x | \vartheta_z \rangle|^2 \right\}_+ |x\rangle\langle x|_{\hat{X}} \otimes |\vartheta_z\rangle\langle \vartheta_z|_S, \quad (53)$$

which satisfies the two constraints of (17). We then have

$$\delta(\mathcal{Q}_Z^\dagger \circ \mathcal{Q}_X, \mathcal{F}) \leq \max_z \sum_x \left\{ \frac{1}{d} - |\langle \varphi_x | \vartheta_z \rangle|^2 \right\}_+. \quad (54)$$

Defining $c_2(X; Z) = 1 - \max_z \sum_x \left\{ \frac{1}{d} - |\langle \varphi_x | \vartheta_z \rangle|^2 \right\}_+$ completes the proof. \square

Note that the proof comes down to how successive measurement of the two observables acts on the system, much like [33].

- Measurement* (Princeton University Press, 1984).
- [3] H. P. Robertson, *Physical Review* **34**, 163 (1929).
 - [4] E. Arthurs and J. L. Kelly, *Bell System Technical Journal* **44**, 725 (1965).
 - [5] C. Y. She and H. Heffner, *Physical Review* **152**, 1103 (1966).
 - [6] E. B. Davies, *Quantum theory of open systems* (Academic Press, London, 1976).
 - [7] S. T. Ali and E. Prugovečki, *Journal of Mathematical Physics* **18**, 219 (1977).
 - [8] E. Prugovečki, *Journal of Physics A: Mathematical and General* **10**, 543 (1977).
 - [9] P. Busch, *International Journal of Theoretical Physics* **24**, 63 (1985).
 - [10] P. Busch, *Physical Review D* **33**, 2253 (1986).
 - [11] E. Arthurs and M. S. Goodman, *Physical Review Letters* **60**, 2447 (1988).
 - [12] H. Martens and W. M. de Muynck, *Physics Letters A* **157**, 441 (1991).
 - [13] S. Ishikawa, *Reports on Mathematical Physics* **29**, 257 (1991).
 - [14] M. G. Raymer, *American Journal of Physics* **62**, 986 (1994).
 - [15] U. Leonhardt, B. Böhmer, and H. Paul, *Optics Communications* **119**, 296 (1995).
 - [16] D. M. Appleby, *International Journal of Theoretical Physics* **37**, 1491 (1998).
 - [17] M. J. W. Hall, *Physical Review A* **69**, 052113 (2004).
 - [18] R. F. Werner, *Quantum Information and Computation* **4**, 546 (2004).
 - [19] M. Ozawa, *Physics Letters A* **320**, 367 (2004).
 - [20] Y. Watanabe, T. Sagawa, and M. Ueda, *Physical Review A* **84**, 042121 (2011).
 - [21] P. Busch, P. Lahti, and R. F. Werner, *Physical Review Letters* **111**, 160405 (2013).
 - [22] P. Busch, P. Lahti, and R. F. Werner, *Physical Review A* **89**, 012129 (2014).
 - [23] P. Busch, P. Lahti, and R. F. Werner, *arXiv:1312.4392 [quant-ph]* (2013).
 - [24] V. B. Braginsky and F. Y. Khalili, *Quantum Measurement* (Cambridge University Press, 1992).
 - [25] H. Martens and W. M. d. Muynck, *Journal of Physics A: Mathematical and General* **25**, 4887 (1992).
 - [26] M. Ozawa, *Physical Review A* **67**, 042105 (2003).
 - [27] Y. Watanabe and M. Ueda, *arXiv:1106.2526 [quant-ph]* (2011).
 - [28] C. Branciard, *Proceedings of the National Academy of Sciences* **110**, 6742 (2013).
 - [29] F. Buscemi, M. Hall, M. Ozawa, and M. M. Wilde, *Phys. Rev. Lett.* **112**, 050401 (2014).
 - [30] A. C. Ipsen, *arXiv:1311.0259 [quant-ph]* (2013).
 - [31] P. J. Coles and F. Furrer, *arXiv:1311.7637 [quant-ph]* (2013).
 - [32] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, *Nature Physics* **6**, 659 (2010).
 - [33] M. Tomamichel and R. Renner, *Physical Review Letters* **106**, 110506 (2011).
 - [34] M. Berta, M. Christandl, F. Furrer, V. B. Scholz, and M. Tomamichel, *arXiv:1308.4527 [quant-ph]* (2013).
 - [35] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, *Nature Communications* **3**, 634 (2012).
 - [36] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B. Scholz, M. Tomamichel, and R. F. Werner, *Physical Review Letters* **109**, 100502 (2012).
 - [37] D. Kretschmann, D. Schlingemann, and R. F. Werner, *Journal of Functional Analysis* **255**, 1889 (2008).
 - [38] D. Kretschmann, D. Schlingemann, and R. Werner, *IEEE Transactions on Information Theory* **54**, 1708 (2008).
 - [39] F. G. Lacerda, J. M. Renes, and R. Renner, *Classical leakage resilience from fault-tolerant quantum computation*, in preparation.
 - [40] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, Lecture notes in physics No. 190 (Springer-Verlag, Berlin, 1983).
 - [41] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
 - [42] W. F. Stinespring, *Proceedings of the American Mathematical Society* **6**, 211 (1955).
 - [43] V. Paulsen, *Completely Bounded Maps and Operator Algebras*, Cambridge Studies in Advanced Mathematics, Vol. 78 (Cambridge University Press, 2003).
 - [44] P. W. Shor and J. Preskill, *Physical Review Letters* **85**, 441 (2000).
 - [45] I. Devetak, *Information Theory*, *IEEE Transactions on* **51**, 44 (2005).
 - [46] J. M. Renes, *Proceedings of the Royal Society A* **467**, 1604 (2011).
 - [47] J. M. Renes, *The Physics of Quantum Information: Complementarity, Uncertainty, and Entanglement*, Habilitation thesis, TU Darmstadt (2012), *arXiv:1212.2379 [quant-ph]*.
 - [48] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, *Nature Physics* **8**, 185 (2012).
 - [49] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, *Physical Review Letters* **109**, 100404 (2012).
 - [50] M. Owari, M. B. Plenio, E. S. Polzik, A. Serafini, and M. M. Wolf, *New Journal of Physics* **10**, 113014 (2008).
 - [51] C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften (Springer-Verlag, Berlin, 2008).
 - [52] A. Y. Kitaev, *Russian Mathematical Surveys* **52**, 1191 (1997).
 - [53] M. F. Sacchi, *Physical Review A* **72**, 014305 (2005).
 - [54] C. W. Helstrom, *Information and Control* **10**, 254 (1967).
 - [55] C. W. Helstrom, *Quantum detection and estimation theory*, Mathematics in Science and Engineering, Vol. 123 (Academic Press, London, 1976).
 - [56] J. Watrous, *Theory of Computing* **5**, 217 (2009).